



Photometry-based visual servoing using light reflexion models.

Christophe Collewet, Eric Marchand

► To cite this version:

Christophe Collewet, Eric Marchand. Photometry-based visual servoing using light reflexion models.. IEEE Int. Conf. on Robotics and Automation, ICRA'09, 2009, Kobe, Japan, Japan. pp.701-706. inria-00436733

HAL Id: inria-00436733

<https://inria.hal.science/inria-00436733>

Submitted on 27 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Photometry-based visual servoing using light reflexion models

Christophe Collewet and Eric Marchand

Abstract—We present in this paper a way to achieve positioning tasks by visual servoing under complex luminance variations. To do that, we use as visual features the luminance of all pixels in the image as we did in our previous work [4]. An important issue of this approach is that it does not rely at all on a any matching nor tracking process, contrary to all the approaches based on geometric visual features (points, straight lines, pose, homography, etc.). However, we consider in this paper a complete illumination model contrary to [4] where the temporal luminance constancy hypothesis was assumed. The main issue of this paper is thus the analytical computation of the interaction matrix related to the luminance from the Phong illumination model. Experimental results on specular objects validate our approach.

I. INTRODUCTION

Visual servoing is a well known technique in robot control that allows to achieve robotic tasks from visual features acquired by a camera [2]. In such an approach, the control law is designed to move a robot so that the current visual features \mathbf{s} , acquired from the current pose \mathbf{r} , will reach the desired features \mathbf{s}^* acquired at the desired pose \mathbf{r}^* . The control law is thus designed to vanish the error $\mathbf{e} = \mathbf{s} - \mathbf{s}^*$. This requires the knowledge of the so-called interaction matrix [6]. This matrix links the time variation of \mathbf{s} to the camera instantaneous velocity \mathbf{v} :

$$\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v} \quad (1)$$

with $\mathbf{v} = (\mathbf{v}, \boldsymbol{\omega})$ where \mathbf{v} and $\boldsymbol{\omega}$ are respectively the linear and angular camera velocities.

However, to deal with complex object appearance, this approach requires robust image processings. Indeed, whatever the visual features \mathbf{s} we choose (points, straight lines, pose, homography, etc.), they are always estimated from visual measurements $\mathbf{m}(\mathbf{r})$ that we have to extract from the initial image but also to match with $\mathbf{m}(\mathbf{r}^*)$ extracted from the desired frame. Moreover, one need also to keep this matching all along the motion thanks to a dedicated tracking algorithm. That means that visual servoing requires *a priori* knowledges about the texture of the scene. Besides, tracking is known to be a complex task as proved by the large amount of works on this subject, especially when occlusions or complex luminance variations occur.

In this paper, we propose a way to remove these hypothesis but also to avoid any image processing, even in the case of specular highlights occurrence. Indeed, no other information than the luminance I (the pure image signal) is involved in the control law. Thus, the visual feature vector \mathbf{s} is nothing

but the image while \mathbf{s}^* is the desired image, the error \mathbf{e} is thus simply the difference between these two images.

Considering the whole image as a feature has previously been considered in [5], [9]. As in our case, these methods did not require any matching process. Nevertheless they differ from our approach in two important points. First, they do not use directly the image intensity. The control is performed from an eigenspace decomposition. This requires the off-line computation of this eigenspace (using a principal component analysis) and then, for each new frame, the projection of the image on this subspace. Second, the interaction matrix related to the eigenspace is not computed analytically but learned during an off-line step. This learning process has two drawbacks: it has to be done for each new object and if the direction of the lighting has been changed. It also requires the acquisition of many images of the scene at various camera positions. Considering an analytical interaction matrix avoids these issues. Pixels intensity has also been considered in [7]. More precisely, a kernel method has been used leading to a high decoupled control law. However, only the translations and the rotation around the optical axis are considered whereas, in our work, the 6 degrees of freedom are controlled. Finally, in [11], an homography-based visual servoing has been proposed since it is used to design the control law. Despite the fact that, as in our case, image intensity is used as the basis of the approach, an important image processing step is necessary to estimate this homography.

The main contribution of this paper is then to analytically compute the interaction matrix related to the luminance. It has been already computed in our previous work [8] under the temporal luminance constancy hypothesis. However, if we want to consider more realistic scenes (i.e. non Lambertian scenes) we have to use a reflection model such as one described in [1], [10], [12]. Indeed, we will use the Phong reflection model [10] to tackle more complex illumination changes. In this way we will be able to take into account that the camera is moving with respect to the light position or that specular highlights occur, without requiring any image processing. It is the contribution of this work.

II. INTERACTION MATRIX RELATED TO THE LUMINANCE

The visual features considered in this paper are the luminance I of each point of the image. We thus have

$$\mathbf{s}(\mathbf{r}) = \mathbf{I}(\mathbf{r}) = (\mathbf{I}_{1\bullet}, \mathbf{I}_{2\bullet}, \dots, \mathbf{I}_{N\bullet}) \quad (2)$$

where $\mathbf{I}_{k\bullet}$ is the k -th line of the image. $\mathbf{I}(\mathbf{r})$ is then a vector of size $N \times M$ where $N \times M$ is the size of the image. As already stated, to derive the computation of the

interaction matrix, we have to consider a more realistic reflection model than the Lambert's one. We consider here the well known Phong model [10] (see Fig. 1) which provides the intensity $I(\mathbf{x})$ at point $\mathbf{x} = (x, y, 1)$, \mathbf{x} being the perspective projection of a physical point \mathbf{P} belonging to the scene:

$$I(\mathbf{x}) = K_s \cos^k \alpha + K_d \cos \theta + K_a. \quad (3)$$

This relation is composed of a diffuse, a specular and an ambient component and assumes a point light source. The scalar k is inversely proportional to the roughness of the object surface; K_s describes the specular component of the lighting; K_d describes the part of the diffuse term which depends on the *albedo* in \mathbf{P} ; K_a is the intensity of ambient lighting in \mathbf{P} . θ is the angle between the normal to the surface \mathbf{n} in \mathbf{P} and the direction of the light source \mathbf{L} ; α is the angle between \mathbf{R} (which is \mathbf{L} mirrored about \mathbf{n}) and the viewing direction \mathbf{V} . Although empirical, this model is widely used thanks to its simplicity, and because it is appropriate for various types of materials, whether they are rough or smooth.

Let us introduce the unit vectors \mathbf{i}, \mathbf{j} and \mathbf{k} corresponding to the axis of the camera frame that will be useful in the remainder of this paper (see Fig. 1)

Considering that \mathbf{R}, \mathbf{V} and \mathbf{L} are normalized, we can rewrite (3) as

$$I(\mathbf{x}) = K_s u_1^k + K_d u_2 + K_a \quad (4)$$

where $u_1 = \mathbf{R}^\top \mathbf{V}$ while we have $u_2 = \mathbf{n}^\top \mathbf{L}$. Note that these vectors are easy to compute, indeed we have

$$\mathbf{V} = -\frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (5)$$

$$\mathbf{R} = 2u_2 \mathbf{n} - \mathbf{L}. \quad (6)$$

In the general case, we will consider the following dependences

$$\begin{cases} \mathbf{V} = \mathbf{V}(\mathbf{x}(t)) \\ \mathbf{n} = \mathbf{n}(\mathbf{x}(t), t) \\ \mathbf{L} = \mathbf{L}(\mathbf{x}(t), t) \\ \mathbf{R} = \mathbf{R}(\mathbf{x}(t), t). \end{cases} \quad (7)$$

From the definition of the interaction matrix given in (1), computing it requires to write the total derivative of (4)

$$\dot{I} = k K_s u_1^{k-1} \dot{u}_1 + K_d \dot{u}_2. \quad (8)$$

However, it is also possible to compute \dot{I} as

$$\dot{I} = \nabla I^\top \dot{\mathbf{x}} + I_t = \nabla I^\top \mathbf{L}_x \mathbf{v} + I_t \quad (9)$$

where we have introduced the interaction matrix \mathbf{L}_x associated to \mathbf{x} and where ∇I_x and ∇I_y are the components along x and y of the spatial gradient ∇I of the image. Note that it is actually the only image processing step necessary to implement our method. Consequently, from (8) and (9), we obtain¹

$$I_t = -\nabla I^\top \mathbf{L}_x \mathbf{v} + k K_s u_1^{k-1} \dot{u}_1 + K_d \dot{u}_2. \quad (10)$$

¹Note that this equation (10) is a new formulation of the classical optical flow constraint equation used in most of optical flow estimation algorithms or SSD based algorithms (such as the well known KLT one).

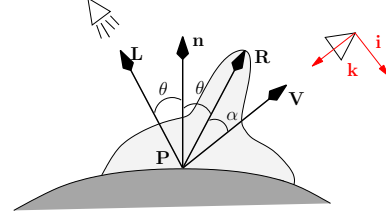


Fig. 1. The Phong illumination model [10].

Thus, by explicitly computing the total time derivative of u_1 and u_2 and writing

$$\dot{u}_1 = \mathbf{L}_1^\top \mathbf{v} \text{ and } \dot{u}_2 = \mathbf{L}_2^\top \mathbf{v}, \quad (11)$$

we obtain the interaction matrix related to the intensity at pixel \mathbf{x} in the general case:

$$\mathbf{L}_I = -\nabla I^\top \mathbf{L}_x + k K_s u_1^{k-1} \mathbf{L}_1^\top + K_d \mathbf{L}_2^\top. \quad (12)$$

Note that we recover here the interaction matrix $-\nabla I^\top \mathbf{L}_x$ associated to the intensity when the temporal luminance constancy hypothesis is assumed (see [8]).

We now compute the vectors \mathbf{L}_1 and \mathbf{L}_2 involved in (11) to explicitly compute \mathbf{L}_I .

A. Computation of \mathbf{L}_2

This computation requires to write

$$\dot{u}_2 = \mathbf{L}^\top \dot{\mathbf{n}} + \mathbf{n}^\top \dot{\mathbf{L}}. \quad (13)$$

In the general case of the time dependences given by (7), (13) becomes

$$u_2 = \mathbf{L}^\top \left(\mathbf{J}^n \dot{\mathbf{x}} + \frac{\partial \mathbf{n}}{\partial t} \right) + \mathbf{n}^\top \left(\mathbf{J}^L \dot{\mathbf{x}} + \frac{\partial \mathbf{L}}{\partial t} \right) \quad (14)$$

$$= (\mathbf{L}^\top \mathbf{J}^n + \mathbf{n}^\top \mathbf{J}^L) \dot{\mathbf{x}} + \mathbf{L}^\top \frac{\partial \mathbf{n}}{\partial t} + \mathbf{n}^\top \frac{\partial \mathbf{L}}{\partial t} \quad (15)$$

where \mathbf{J}^n and \mathbf{J}^L express respectively the Jacobian matrices related to \mathbf{n} and \mathbf{L} with respect to \mathbf{x} .

However, if we want to express $\partial \mathbf{n} / \partial t$ and $\partial \mathbf{L} / \partial t$ in function of the camera velocity \mathbf{v} , an apriori knowledge about the scene is required. Is the object moving or not? Is the light moving or not wrt the object? In this paper we consider two particular cases, although this approach is, of course, general.

- *Light source motionless with respect to the object frame.* To compute $\partial \mathbf{n} / \partial t$, we introduce the matrix ${}^c \mathbf{R}_o$ which describes the rotation between the camera and the object frames such that $\mathbf{n} = {}^c \mathbf{R}_o {}^o \mathbf{n}$ where ${}^o \mathbf{n}$ expresses \mathbf{n} in the object frame. Therefore,

$$\frac{\partial \mathbf{n}}{\partial t} = {}^c \dot{\mathbf{R}}_o {}^o \mathbf{n} = {}^c \dot{\mathbf{R}}_o {}^c \mathbf{R}_o^\top \mathbf{n} = -\boldsymbol{\omega} \times \mathbf{n}. \quad (16)$$

Note that we similarly have

$$\frac{\partial \mathbf{L}}{\partial t} = -\boldsymbol{\omega} \times \mathbf{L}. \quad (17)$$

In that case, it is straightforward to show that

$$\begin{aligned}\frac{\partial u_2}{\partial t} &= \mathbf{L}^\top \frac{\partial \mathbf{n}}{\partial t} + \mathbf{n}^\top \frac{\partial \mathbf{L}}{\partial t} \\ &= 0\end{aligned}\quad (18)$$

which directly gives from (15)

$$\mathbf{L}_2^\top = (\mathbf{L}^\top \mathbf{J}^n + \mathbf{n}^\top \mathbf{J}^L) \mathbf{L}_x \quad (20)$$

where we have introduced the interaction matrix of \mathbf{x} .

• *Light source mounted on the camera.* We thus have $\mathbf{L} = -\mathbf{k}$. In that case, the equations become simpler since we have $\mathbf{J}^L = \mathbf{0}$ and $\partial \mathbf{L} / \partial t = \mathbf{0}$.

Let us rewrite u_2 from (15) where we have introduced \mathbf{L}_x and (16)

$$u_2 = -(\mathbf{k}^\top \mathbf{J}^n) \mathbf{L}_x \mathbf{v} - \mathbf{k}^\top (\mathbf{n} \times \boldsymbol{\omega}) \quad (21)$$

$$= -\nabla n_z^\top \mathbf{L}_x \mathbf{v} + (\mathbf{n} \times \mathbf{k})^\top \boldsymbol{\omega}. \quad (22)$$

Therefore, by introducing the following vector

$$\mathbf{L}_4^\top = \begin{pmatrix} 0 & 0 & 0 & (\mathbf{n} \times \mathbf{k})^\top \mathbf{i} & (\mathbf{n} \times \mathbf{k})^\top \mathbf{j} & 0 \end{pmatrix}, \quad (23)$$

\mathbf{L}_2^\top expresses as follows

$$\mathbf{L}_2^\top = -\nabla n_z^\top \mathbf{L}_x + \mathbf{L}_4^\top. \quad (24)$$

B. Computation of \mathbf{L}_1

Let us recall that $u_1 = \mathbf{R}^\top \mathbf{V}$ leading, by considering the time dependences given in (7), to

$$\dot{u}_1 = \mathbf{V}^\top \dot{\mathbf{R}} + \mathbf{R}^\top \dot{\mathbf{V}} \quad (25)$$

which gives

$$u_1 = \mathbf{V}^\top \left(\mathbf{J}^R \dot{\mathbf{x}} + \frac{\partial \mathbf{R}}{\partial t} \right) + \mathbf{R}^\top \mathbf{J}^V \dot{\mathbf{x}} \quad (26)$$

$$= (\mathbf{V}^\top \mathbf{J}^R + \mathbf{R}^\top \mathbf{J}^V) \dot{\mathbf{x}} + \mathbf{V}^\top \frac{\partial \mathbf{R}}{\partial t} \quad (27)$$

where \mathbf{J}^R and \mathbf{J}^V express respectively the Jacobian matrices related to \mathbf{R} and \mathbf{V} with respect to \mathbf{x} . In addition, since \mathbf{R} is a function of \mathbf{L} and \mathbf{n} (see (6)), we have

$$\frac{\partial \mathbf{R}}{\partial t} = 2 \left(\frac{\partial u_2}{\partial t} \mathbf{n} + u_2 \frac{\partial \mathbf{n}}{\partial t} \right) - \frac{\partial \mathbf{L}}{\partial t} \quad (28)$$

and after some manipulations detailed in [3]

$$\mathbf{J}^R = 2\mathbf{n}(\mathbf{L}^\top \mathbf{J}^n + \mathbf{n}^\top \mathbf{J}^L) + 2u_2 \mathbf{J}^n - \mathbf{J}^L. \quad (29)$$

The computation of \mathbf{J}^V is also very simple, see [3].

At this step, as we did for the computation of \mathbf{L}_2 , we consider the cases when the light source is motionless with respect to the object frame or when the light source is mounted on the camera.

• *Light source motionless with respect to the object frame.* From (18), (16) and (17), (28) becomes

$$\frac{\partial \mathbf{R}}{\partial t} = 2u_2(\mathbf{n} \times \boldsymbol{\omega}) - (\mathbf{L} \times \boldsymbol{\omega}) = \mathbf{R} \times \boldsymbol{\omega} \quad (30)$$

and by introducing the vector \mathbf{L}_3^\top such as

$$\mathbf{V}^\top \frac{\partial \mathbf{R}}{\partial t} = \mathbf{L}_3^\top \mathbf{v} \quad (31)$$

$$= \begin{pmatrix} 0 & 0 & 0 & \mathbf{W}^\top \mathbf{i} & \mathbf{W}^\top \mathbf{j} & \mathbf{W}^\top \mathbf{k} \end{pmatrix} \mathbf{v} \quad (32)$$

with $\mathbf{W} = \mathbf{V} \times \mathbf{R}$, we obtain \mathbf{L}_1^\top from (27) by introducing \mathbf{L}_x

$$\mathbf{L}_1^\top = (\mathbf{V}^\top \mathbf{J}^R + \mathbf{R}^\top \mathbf{J}^V) \mathbf{L}_x + \mathbf{L}_3^\top. \quad (33)$$

• *Light source mounted on the camera.* Recall that we have in this case $\mathbf{L} = -\mathbf{k}$ leading to $\mathbf{J}^L = \mathbf{0}$ and $\partial \mathbf{L} / \partial t = \mathbf{0}$.

To rewrite u_1 given in (27), we first compute (28) using $u_2 = -\mathbf{k}^\top \mathbf{n}$

$$\frac{\partial \mathbf{R}}{\partial t} = -2 \left(\mathbf{k}^\top \frac{\partial \mathbf{n}}{\partial t} \mathbf{n} + \mathbf{k}^\top \mathbf{n} \frac{\partial \mathbf{n}}{\partial t} \right) \quad (34)$$

leading by using (16) to

$$\mathbf{V}^\top \frac{\partial \mathbf{R}}{\partial t} = -2(\mathbf{k}^\top (\mathbf{n} \times \boldsymbol{\omega}) \mathbf{V}^\top \mathbf{n} + \mathbf{k}^\top \mathbf{n} \mathbf{V}^\top (\mathbf{n} \times \boldsymbol{\omega})) \quad (35)$$

$$= -2(\mathbf{n}^\top \mathbf{V} (\mathbf{k} \times \mathbf{n})^\top \boldsymbol{\omega} + \mathbf{k}^\top \mathbf{n} (\mathbf{V} \times \mathbf{n})^\top \boldsymbol{\omega}). \quad (36)$$

Consequently, by introducing the vector \mathbf{L}_5^\top such as

$$\mathbf{V}^\top \frac{\partial \mathbf{R}}{\partial t} = \mathbf{L}_5^\top \mathbf{v} \quad (37)$$

$$= \begin{pmatrix} 0 & 0 & 0 & \mathbf{L}_{5x} & \mathbf{L}_{5y} & \mathbf{L}_{5z} \end{pmatrix} \mathbf{v} \quad (38)$$

with

$$\begin{cases} \mathbf{L}_{5x} &= 2(\mathbf{n}^\top \mathbf{V} (\mathbf{n} \times \mathbf{k})^\top + \mathbf{k}^\top \mathbf{n} (\mathbf{n} \times \mathbf{V})^\top) \mathbf{i} \\ \mathbf{L}_{5y} &= 2(\mathbf{n}^\top \mathbf{V} (\mathbf{n} \times \mathbf{k})^\top + \mathbf{k}^\top \mathbf{n} (\mathbf{n} \times \mathbf{V})^\top) \mathbf{j} \\ \mathbf{L}_{5z} &= 2\mathbf{k}^\top \mathbf{n} (\mathbf{n} \times \mathbf{V})^\top \mathbf{k} \end{cases} \quad (39)$$

we obtain from (27)

$$\mathbf{L}_1^\top = (\mathbf{V}^\top \mathbf{J}^R + \mathbf{R}^\top \mathbf{J}^V) \mathbf{L}_x + \mathbf{L}_5^\top. \quad (40)$$

III. INTERACTION MATRIX IN SOME PARTICULAR CASES

In this section, we consider some particular cases very useful in the industrial vision domain. Indeed, in such applications the lighting position is well known wrt the object. We consider also the case of a light-ring mounted on the camera which is also very useful in this application domain.

Moreover, as usually done in visual servoing, the interaction matrix is computed at the desired position [2], that is $\mathbf{n} = -\mathbf{k}$. In addition, since \mathbf{L}_1 and \mathbf{L}_2 depends on \mathbf{J}^n , the shape of the object being observed has to be known to obtain \mathbf{J}^n . We consider here the simple case of a planar object but this computation can be easily done in some other cases like the cases of a cylinder or a sphere for example.

A. Light source motionless with respect to the object frame and located at infinity

This case is depicted on figure 2. Since a planar object is considered, \mathbf{n} does no longer depend on \mathbf{x} , then $\mathbf{J}^n = \mathbf{0}$. Similarly, since the light source is at infinity, \mathbf{L} does not depend on \mathbf{x} and thus $\mathbf{J}^L = \mathbf{0}$. In addition, since the angle between \mathbf{n} and \mathbf{L} is constant, $u_2 = \mathbf{n}^\top \mathbf{L}$ is constant.

Consequently, it is easy to show from (20) that $\mathbf{L}_2^\top = \mathbf{0}$.

For \mathbf{L}_1^\top , since $\mathbf{J}^n = \mathbf{J}^L = \mathbf{0}$, we have also $\mathbf{J}^R = \mathbf{0}$ (see (29)), leading from (33) to

$$\mathbf{L}_1^\top = \mathbf{R}^\top \mathbf{J}^V \mathbf{L}_x + \mathbf{L}_3^\top. \quad (41)$$

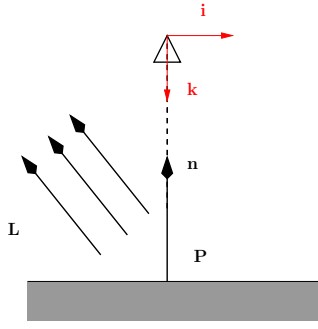


Fig. 2. Light source located at infinity and camera and object planes parallel.

Thus, \mathbf{L}_1^\top can be easily computed. All computations done, we obtain

$$\mathbf{L}_1^\top = \frac{1}{Z\|\mathbf{x}\|^3} \begin{pmatrix} L_{1x} & L_{1y} & L_{1z} & 0 & 0 & 0 \end{pmatrix} \quad (42)$$

with

$$\begin{cases} L_{1x} = (1 + y^2) R_x - xyR_y - xR_z \\ L_{1y} = -xyR_x + (1 + x^2) R_y - yR_z \\ L_{1z} = -xR_x - yR_y + (x^2 + y^2) R_z \end{cases} \quad (43)$$

where R_x, R_y, R_z are the components of \mathbf{R} that can be easily evaluated (\mathbf{R} is \mathbf{L} mirrored about \mathbf{n}).

B. Light source mounted on the camera

This case is depicted on the figure 3. Since $\mathbf{J}^n = \mathbf{0}$ and $\mathbf{n} = -\mathbf{k}$, from (23) and (24) it is straightforward to show that $\mathbf{L}_2^\top = \mathbf{0}$. Besides, since $\mathbf{n} = -\mathbf{k}$ and $\mathbf{L} = -\mathbf{k}$, we have $\mathbf{R} = -\mathbf{k}$. We also have $\mathbf{J}^R = \mathbf{0}$ since $\mathbf{J}^n = \mathbf{J}^L = \mathbf{0}$. Consequently \mathbf{L}_1^\top becomes from (40)

$$\mathbf{L}_1^\top = -\mathbf{k}^\top \mathbf{J}^V \mathbf{L}_x + \mathbf{L}_5^\top \quad (44)$$

while \mathbf{L}_5^\top becomes from (39)

$$\mathbf{L}_5^\top = \begin{pmatrix} 0 & 0 & 0 & -2\mathbf{V}^\top \mathbf{j} & -2\mathbf{V}^\top \mathbf{i} & 0 \end{pmatrix}. \quad (45)$$

Using explicitly \mathbf{V} , \mathbf{J}^V and \mathbf{L}_x , we simply obtain

$$\mathbf{L}_1^\top = \frac{1}{\|\mathbf{x}\|} \begin{pmatrix} \frac{x}{Z} & \frac{y}{Z} & -\frac{x^2 + y^2}{Z} & y & -x & 0 \end{pmatrix} \quad (46)$$

where $\bar{Z} = Z\|\mathbf{x}\|^2$.

IV. CONTROL LAW

Once the interaction matrix is known, the control law can be designed. We chose the same as we proposed in [4]:

$$\mathbf{v} = -\lambda (\mathbf{H}^* + \mu \text{diag}(\mathbf{H}^*))^{-1} \mathbf{L}_{I^*}^\top (\mathbf{I}(\mathbf{r}) - \mathbf{I}(\mathbf{r}^*)) \quad (47)$$

where $\mathbf{H}^* = \mathbf{L}_{I^*}^\top \mathbf{L}_{I^*}$ and μ a positive scalar.

Note that the interaction matrix \mathbf{L}_{I^*} is related to the luminance $\mathbf{I}^*(\mathbf{r})$ at each point of the image acquired at the desired position \mathbf{r}^* as it is usually the case in visual servoing. Moreover, we will consider that \mathbf{r}^* is such that the camera and the object planes are parallel so that the matrices computed in section III can be used. In addition, we consider that the depths are constant and equal to those estimated at the desired pose ($\forall \mathbf{x}, Z(\mathbf{x}) = Z^* \approx 70$ cm).

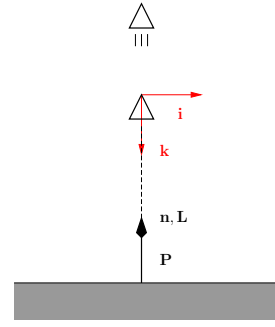


Fig. 3. Light source mounted on the camera for a planar object when the camera and the object planes are parallel.

In addition, let us point out that (47) looks like a Levenberg-Marquardt based control law. However, the way to tune μ is very different. All the details are given in [4]. Indeed, the classical control failed, either because they diverged or because they led to unsuitable 3D motion.

V. EXPERIMENTAL RESULTS

The experiments described below have been carried out on a 6 d.o.f gantry robot. In our previous paper [4], we considered a temporal luminance constancy model of the environment and it has been shown that this approach was quite robust to modeling errors, depth uncertainty (that is able to deal with non planar objects), and partial occlusions. Nevertheless, in this paper we have extended the model of illumination in order to consider more complex models that includes specularities. We will show that, considering the extended model, the behavior of the robot is better than in the previous case. Let us point out that the images are acquired at 66Hz using an IEEE 1394 camera with a resolution of 320x240. The size of the vector \mathbf{s} is then 76800.

A. Light source motionless with respect to the object frame and located at infinity

We carried out the experiments on various scenes and on various initial positions. In this set of experiments a unique directional light has been added to the scene with a 45° rotation around the vector \mathbf{j} of the camera frame (see Fig. 2 for an illustration). This light produces an important specularity on the scene and then on the image (as can be shown on Figures 5 and 6).

For this first experiment the initial positioning error was $\Delta \mathbf{r} = (-23 \text{ mm}, -201 \text{ mm}, 93 \text{ mm}, -17.7^\circ, 1.5^\circ, -4.8^\circ)$. First, let us note that in both cases (temporal luminance constancy model or the new proposed model) the robot converges toward the desired position (see Fig. 4). Nevertheless the convergence rate when considering explicitly specularities is always faster and smoother (this is due to a better estimation of the gradient of the cost function $\|\mathbf{e}\|^2$ that is used in the minimization process).

A similar experiment, although with a larger initial error, is shown on Figure 5. Indeed, the initial positioning error is $\Delta \mathbf{r} = (-24 \text{ mm}, -176 \text{ mm}, 86 \text{ mm}, -13.75^\circ, -6.76^\circ, -30.53^\circ)$. A similar behavior can be observed. Images shows the current images (images 0, 300 and 900) and the corresponding

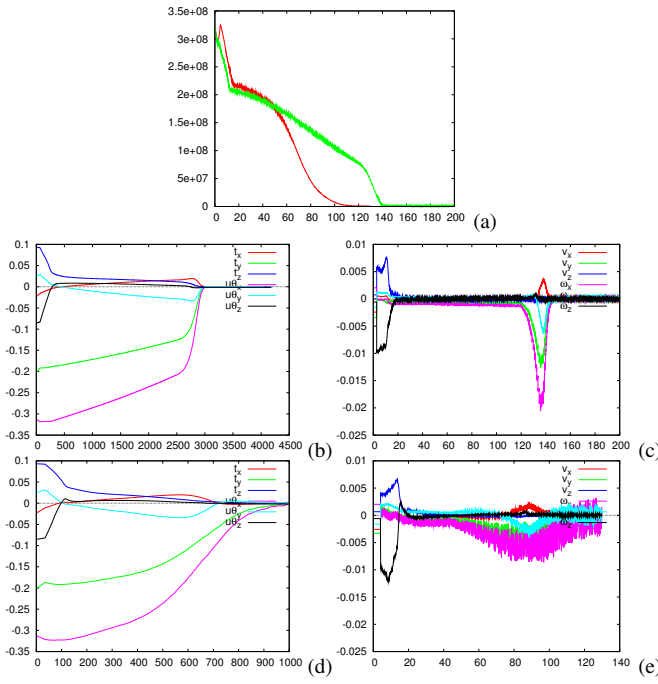


Fig. 4. Positioning task with a light source motionless with respect to the object frame and located at “infinity” (x axis in second for (a), (c) and (e), frame number for (b) and (d)). (a) cost function with a temporal luminance constancy model (green) and complete model (red), (b) positioning error (in m, rad) and (c) camera velocity for a temporal luminance constancy model, (d) positioning error (in m, rad) and (e) camera velocity for the complete model.

error $\mathbf{I} - \mathbf{I}^*$. Note that the specularity can be mainly seen near the head of the football player but as can be seen on the right images, it is in fact located all around the image. The final positioning error is very low since we have $\Delta \mathbf{r} = (1 \text{ mm}, 0.04 \text{ mm}, 0 \text{ mm}, 0.04^\circ, 0.08^\circ, 0.1^\circ)$. It is because $\mathbf{I} - \mathbf{I}^*$ is very sensitive to the pose \mathbf{r} .

The third experiment (see Fig. 6) describes the same experiment but with another and less textured scene. Here again a large specularity can be seen in the image. Despite these difficulties, the camera converges smoothly toward the desired position. The error $\mathbf{I} - \mathbf{I}^*$ is displayed on the top of Figure 6. The final positioning error is here again very low, we have $\Delta \mathbf{r} = (0 \text{ mm}, 0.02 \text{ mm}, 0.02 \text{ mm}, 0.01^\circ, 0.01^\circ, 0.05^\circ)$.

B. Light source mounted on the camera

In the second set of experiments a light-ring is located around the camera lens (see Fig. 7). Therefore the light direction is aligned with the camera optical axis as described on Figure 3 and thus it is moving with respect to the scene. This is the unique light in the scene. Note also that the initial positioning error is unchanged.

In that case the specularities (see Fig. 8) is maximum (the light ring is only 70 cm above the scene) and consequently its motion in the image is important (for example the specularity can be seen at the bottom of the image in the first image whereas it has moved to the middle at the end of the positioning task. It also saturates the image meaning that no information is available around the specularity. Here again the behavior of the robot is better (that is faster and smoother)

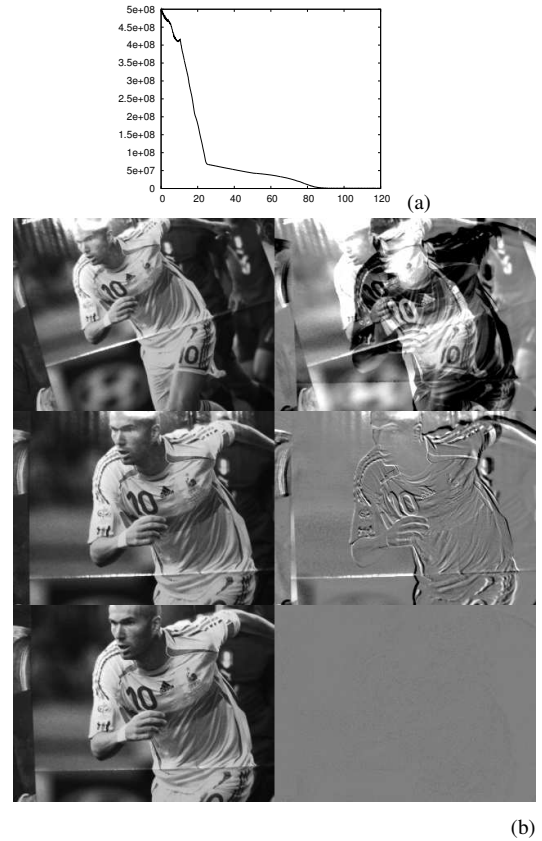


Fig. 5. Positioning task with a light source motionless with respect to the object frame and located at “infinity”. (a) cost function (x axis in second) (b) images acquired during the positioning task (left) and error image $\mathbf{I} - \mathbf{I}^*$ (right).

when the complete model is considered.

VI. CONCLUSION

In this paper we have shown that it is possible to use the luminance of all the pixels in an image as visual features in visual servoing, even when considering complex illumination variations. To do that, we have removed the classical temporal luminance constancy hypothesis and proposed a general framework to compute the variation of the image intensity due to camera and light motions. We have also provided the complete model when the light source is mounted on the camera or located at infinity. Experimental results concerning positioning tasks have been presented on various scenes and on various initial positions which validate our approach. They have always led to very low positioning errors.

REFERENCES

- [1] J. Blinn. Models of light reflection for computer synthesized pictures. In *ACM Conf. on Computer graphics and interactive techniques, SIGGRAPH'77*, pages 192–198, San Jose, California, 1977.
- [2] F. Chaumette and S. Hutchinson. Visual servoing and visual tracking. In B. Siciliano and O. Khatib, editors, *Handbook of Robotics*, chapter 24, pages 563–583. Springer, 2008.
- [3] C. Collewet and E. Marchand. Photometric visual servoing. Technical Report No. 6631, INRIA, September 2008.
- [4] C. Collewet, E. Marchand, and F. Chaumette. Visual servoing set free from image processing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'08*, pages 81–86, Pasadena, California, May 2008.

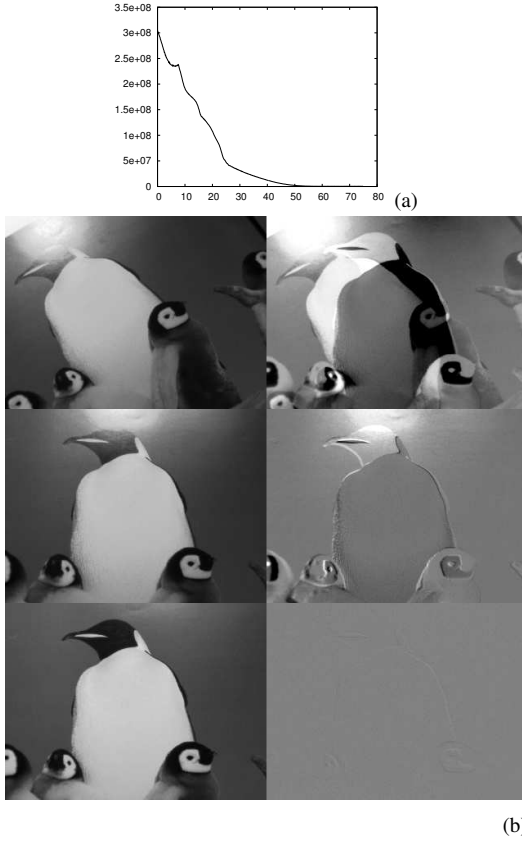


Fig. 6. Positioning task with a light source motionless with respect to the object frame and located at “infinity”. (a) cost function (x axis in second) (b) images acquired during the positioning task (left) and error image $\mathbf{I} - \mathbf{I}^*$ (right).

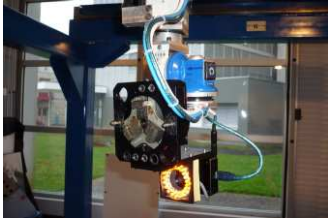


Fig. 7. Camera and light-ring mounted on the robot end-effector.

- [5] K. Deguchi. A direct interpretation of dynamic images with camera and object motions for vision guided robot control. *Int. Journal of Computer Vision*, 37(1):7–20, June 2000.
- [6] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.
- [7] V. Kallem, M. Dewan, J.P. Swensen, G.D. Hager, and N.J. Cowan. Kernel-based visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and System, IROS'07*, San Diego, USA, October 2007.
- [8] E. Marchand. Control camera and light source positions using image gradient information. In *IEEE Int. Conf. on Robotics and Automation, ICRA'07*, Roma, Italy, April 2007.
- [9] S.K. Nayar, S.A. Nene, and H. Murase. Subspace methods for robot vision. *IEEE Trans. on Robotics*, 12(5):750–758, October 1996.
- [10] B.T. Phong. Illumination for computer generated pictures. *Communication of the ACM*, 18(6):311–317, June 1975.
- [11] G. Silveira and E. Malis. Direct visual servoing with respect to rigid objects. In *IEEE/RSJ Int. Conf. on Intelligent Robots and System, IROS'07*, pages 1963–1968, San Diego, USA, October 2007.
- [12] K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.*, 57:1105–1114, 1967.

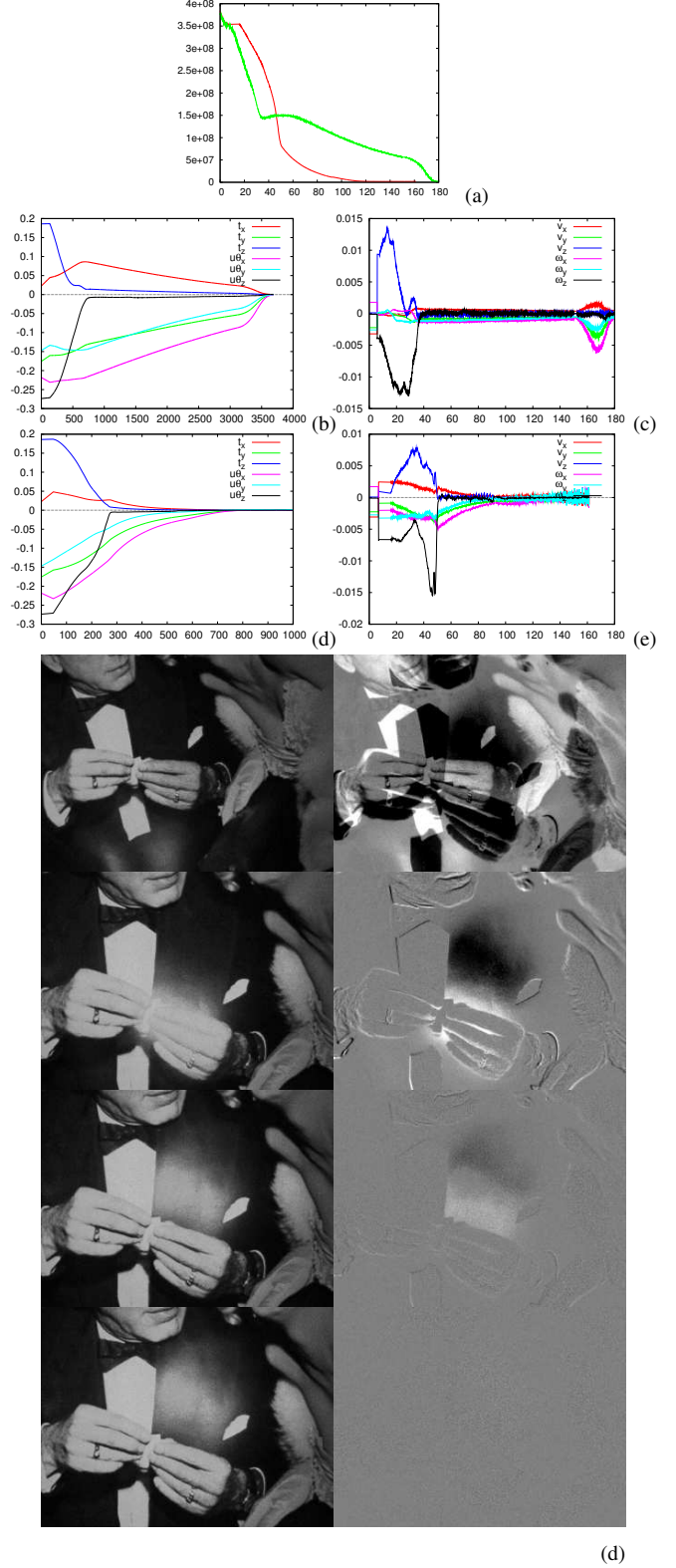


Fig. 8. Positioning task with the light source mounted on the camera. (a) cost function with a temporal luminance constancy model (green) and complete model (red), (b) positioning error (in m, rad) and camera velocity for a temporal luminance constancy model, (c) positioning error (in m, rad) and camera velocity for the complete model. (d) images acquired during the positioning task (left) and error image $\mathbf{I} - \mathbf{I}^*$ (right).